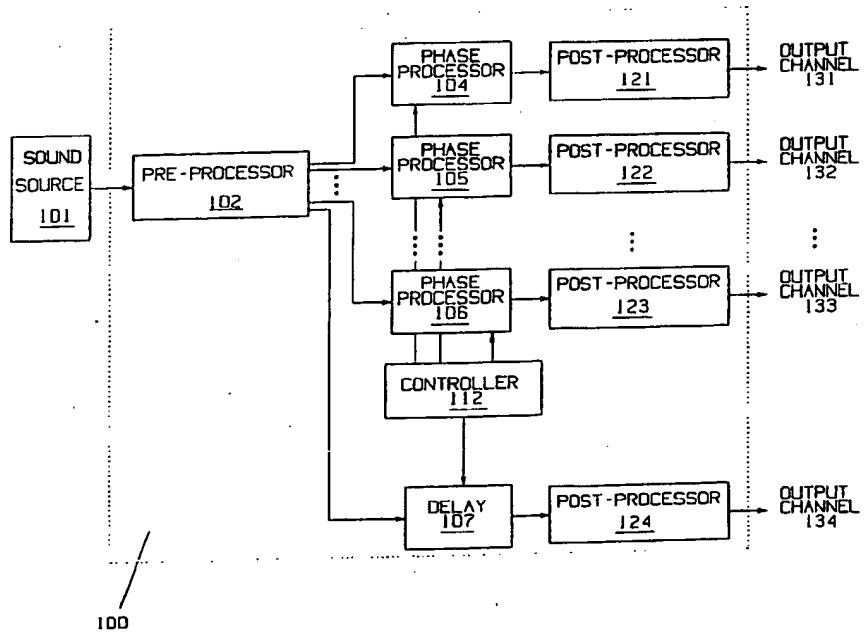




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : H04S 1/00		A1	(11) International Publication Number: WO 91/20165
			(43) International Publication Date: 26 December 1991 (26.12.91)
<p>(21) International Application Number: PCT/US91/04166</p> <p>(22) International Filing Date: 11 June 1991 (11.06.91)</p> <p>(30) Priority data: 538,548 15 June 1990 (15.06.90) US</p> <p>(71) Applicant: AURIS CORP. [US/US]; 1801 Maple Avenue, Evanston, IL 60621 (US).</p> <p>(72) Inventors: WILDE, Martin, D. ; 1516 W. Thorndale, #3E, Chicago, IL 60660 (US). MARTENS, William, M. ; 807 Church Street, #503, Evanston, IL 60201 (US). KENDALL, Gary, S. ; 2111 Madison Place, Evanston, IL 60202 (US).</p> <p>(74) Agents: WARD, Calvin, B. et al.; McCubbrey, Bartels, Meyer & Ward, Suite 2700, One Post Street, San Francisco, CA 94104-5231 (US).</p>		<p>(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).</p> <p>Published <i>With international search report.</i></p>	

(54) Title: IMPROVED AUDIO PROCESSING SYSTEM AND RECORDINGS MADE THEREBY



(57) Abstract

An audio processing system is disclosed in which a plurality of output signals (131, 132, 133, 134) are generated from a sound signal by processing the sound signal (104-107, 121-124) to produce a plurality of signals which have the same intensity in any given frequency band as that of the sound signal. However, the signals in question have altered phase relationships with respect to the sound signal.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

-1-

IMPROVED AUDIO PROCESSING SYSTEM AND RECORDINGS
MADE THEREBY

5

Background of the Invention

10

The present invention relates to acoustical processing systems and more particularly to processing systems in which the acoustical output from a single sound source is processed to produce a plurality of channels.

15
10

Consider a sound processing system which takes a single channel sound source and produces two channels therefrom. For example, the processing system could be a stereophonic system in which the sound source is sensed by two microphones whose outputs are then processed to produce left and right channels for eventual playback on left and right speakers or headphones. Alternatively, the output of a single microphone could be electronically processed to produce the left and right channels. Such a system is described in U.S. Patent 3,670,106.

20
25
30

In the case of stereophonic systems, the goal of the processing system is to create the illusion of a sound source of a predetermined size located at a specific position relative to the speakers. The perceived locations of the various sound sources generated by the stereophonic signals create for the listener what is known as an acoustic image, i.e., a map of the imaginary physical locations of these sound sources. The apparent location of the sound source is largely determined by the difference in arrival time and the intensity of the relevant component signals generated in the left and right speakers.

35

In prior art stereophonic sound systems, the illusion of a sound source of any specific size is difficult to generate in such systems. Some prior art systems utilize reverberation to broaden the sound image. Others utilize 180 degree phase shifts.

35

Shimada (U.S. Patent 3,892,624) and Doi, et al. (U.S. Patent 4,069,394) describe a stereophonic reproduction system in which portions of the input signals are scaled by a constant, k, and cross-fed in 180-degree out-of-phase relationships. That is, given left and right input signals $a_l(t)$ and $a_r(t)$, left and

-2-

right output signals $L = a_L(t) - ka_r(t)$ and $R = a_r(t) - ka_L(t)$ are generated. When L and R are presented over two loudspeakers, a listener located between the

5

loudspeakers perceives a broadened sound image.

These types of systems are problematic in that they often alter the timbral quality of the program material. The summation of the signals used to provide the output signals results in constructive and destructive interference. This interference alters the perceived timbre of the sound. In addition, the acoustical images created often appear broken, and the effects are highly dependent on the listener's location relative to the loudspeakers. The magnitude of these problems depends critically upon the program material; hence, it is impossible to compensate for the distortions through further processing of the resulting signals. As a result, listeners at different locations hear quite different effects in timbre, image width, and image location.

In addition, these systems suffer from two other problems. First, the apparent distance of the sound source is limited to locations on a line between the speakers. For example, the illusion of a sound source located between the speakers and the listener can not be produced without utilizing additional speakers closer to the listener.

Second, the perceived location of the sound source depends critically on the location of the listener relative to the speakers. Thus, if a particular signal component is fed to both speakers with no relative delay and the same signal amplitude, the component of the acoustic image created by that signal will appear to be located on a line centered between the two speakers. If that signal component arrives fractionally earlier from the left speaker than from the right and/or the intensity of the component from the left speaker is greater than that from the right speaker, its image component will appear to be located left of center. The apparent locations of a set of such image components makes up the composite acoustic image perceived by the listener.

In typical listening environments such as living rooms or theaters, most listeners are located nearer to one loudspeaker than to the other(s). For the purposes of the following discussion, it will be assumed that the acoustic image is being produced by only two loudspeakers. If the listener moves nearer to one

25

30

35

-3-

loudspeaker, the sound from that speaker is more intense and reaches the listener ahead of the sound generated at the same time in the other speaker. Hence, moving the listener closer to one speaker is equivalent to introducing an intensity loss and time delay into the material being reproduced in the other speaker. When very similar material is reproduced by two or more loudspeakers, listeners report that the sound images they perceive are either shifted toward the location of the nearest loudspeaker or almost entirely located in the nearest loudspeaker, depending upon the delay in question.

It should be noted that when a listener moves nearer to one speaker, both the intensity of the sound and the time delay are affected. It has been shown that the arrival time difference has a more pronounced and important influence than does the intensity difference.

If the time delay is less than approximately 1.0 msec., listeners describe hearing a single sound image located between the speakers, but shifted toward the closer speaker. This effect is referred to as image shift. If the time delay is greater than approximately 1.0 msec but less than an upper limit discussed below, the listener perceives a single sound image that is located at the closer loudspeaker. The traditional explanation for this phenomenon is that the listener's auditory system has attempted to suppress the delayed signal. This phenomenon is often referred to as the precedence effect, the Haas effect, or the law of the first wavefront. In the following discussion, the effect will be referred to as the precedence effect.

There is an upper limit to the time delay at which the precedence effect operates. At time delays greater than this limit, the delayed sound is heard. The exact magnitude of this upper limit depends upon the qualities of the sound source. The precedence effect is more pronounced for transient sound sources such as struck or plucked musical instruments than it is for continuous sound sources such as blown or bowed musical instruments. The upper limit is found experimentally to vary from 8 to 70 msec with a typical limit being about 15 msec.

When the precedence effect releases, listeners report that the sound image is located in two loudspeakers. When the loudspeakers are separated by a suffi-

ciently great distance, listeners report hearing two sound images, one of which being echo-like. As the time delay from the difference in distances to the two loudspeakers increases further, the intensity difference also increases significantly. When the intensity difference is approximately 15 dB, the more distant loudspeaker becomes difficult to hear. At this point, listeners report that the sound image is located in one loudspeaker.

A further example of a processing system in which a single sound source is processed for reproduction through a number of loudspeakers is a public address system. In such systems, a monophonic signal is reproduced through a plurality of loudspeakers to provide a sound field which covers a large area. These systems suffer from problems of a different type. In those areas in which the acoustical signals produced by different loudspeakers overlap, constructive and destructive interference occurs. The particular frequencies at which these different interference patterns occur is determined by the distance from each of the speakers to the location of the listener. Hence, the sound field at every point in the room will appear to be filtered by a set of frequency filters whose pass-band frequencies depend on the location relative to the speakers. This is equivalent to timbral shifting the original material. Such added coloration is undesirable, since it reduces the intelligibility of the material being broadcast as well as altering the fidelity of the reproduction.

This problem is not limited to monophonic public address systems. Stereophonic systems designed to fill large halls with acoustical sound fields often suffer from this effect.

Broadly, it is an object of the present invention to provide an improved audio processing and reproduction system.

It is yet another object of the present invention to provide a stereophonic system in which the acoustic images are less dependent on the location of the listener relative to the speakers than are the images produced by prior art systems.

It is a further object of the present invention to provide a stereophonic system which provides the illusion of a sound source located between the speak-

-5-

ers and the listener.

5 It is yet another object of the present invention to provide a sound reproduction system for filling a large room with a sound field such that listeners in different parts of the room perceive the same sound field.

10 It is a still further object of the present invention to provide a sound reproduction system which allows the user to control the apparent width and distance of the sound source without adding reverberation or timbre changes.

15 It is yet another object of the present invention to provide a sound reproduction system which allows the user to control the apparent width and distance of the sound source with a minimum of two loudspeakers.

20 It is a still further object of the present invention to provide a sound processing system which allows the apparent width and location of the acoustical image generated thereby to be carried without introducing timbral shifts or causing the image to appear broken.

These and other objects of the present invention will become apparent to those skilled in the art from the following detailed description of the invention and the accompanying drawings.

25 Brief Description of the Drawings

Figure 1 is a block diagram of an audio processing system according to the present invention.

30 Figure 2 is a block diagram of one embodiment of a phase processor according to the present invention.

Summary of the Invention

35 The present invention comprises an apparatus for audio processing, a method of audio procession and a recording made by said method. An audio processing system according to the present invention generates a plurality of

-6-

output signals from a sound source input signal. The system comprises circuitry for receiving the sound input signal and for generating a plurality of channel signals therefrom. One of said channel signals comprises a signal which is substantially equal to the sum of M band-limited signals, the ith said band-limited signal having an amplitude substantially equal to that of said input signal in a predetermined frequency range $f_i \pm \delta f_i$ and a phase which differs from the phase of said input signal in said predetermined frequency range by an amount ϕ_i , i running from 1 to M, wherein $M > 2$ and ϕ_i is chosen between $P - \delta P$ and $P + \delta P$, wherein ϕ_i is a rapidly varying function of i.

Detailed Description of the Invention

For the purpose of the following discussion, it will be assumed that the present invention operates on a single input signal to produce two output signals. The output signals may be channels or may be combined with other material to produce the final channels. The manner in which the present invention would operate to produce more than two output signals will be explained in more detail below.

The present invention provides its beneficial effects by altering the cross-correlation of the output signals while minimizing any timbral shifts between the input signal and the output signals.

The cross-correlation of two signals, $y_1(t)$ and $y_2(t)$, is typically measured in terms of a cross-correlation measure which is defined to be the extreme value of the cross-correlation function $\Omega(x)$, where

$$\Omega(x) = \lim_{T \rightarrow \infty} \frac{1}{(2T)} \int y_1(t)y_2(t+x) dt \quad (1)$$

The cross-correlation measure has a maximum possible value of 1 and a minimum possible value of -1. As will be made clear in the following discussion, it is also important to consider simultaneously both the positive and the negative peaks of the cross-correlation function.

The manner in which the apparatus of the present invention operates may

-7-

be most easily understood with reference to Figures 1 and 2. Figure 1 is block diagram of an audio processing system 100 according to the present invention. Audio processing system receives an input signal from a sound source 100 and produces channels on a plurality of output channels of which output channels 131-134 are exemplary. The input signal from sound source 101 may be in the form of an electric signal or sound waves.

The signal input to audio processor 100 is processed by pre-processor 102 to form a plurality of signals which, after further processing, are incorporated into the signals on the output channels. For the purpose of this discussion, the number of output channels will be denoted by N_{out} . If the input signal is in the form of sound waves, pre-processor 102 includes one or more microphones to convert the sound waves to electrical signals. Pre-processor 102 may be as simple as an electrical junction for dividing the input signal into N_{out} signals.

N_{out} -1 of the signals generated by pre-processor 102 are input to phase-processors of which phase-processors 104-106 are exemplary. The remaining signal may either be input to a phase-processor or to a delay circuit 107.

The manner in which the phase-processing circuit operates may be most easily understood with reference to Figure 2 which is a block diagram of a phase-processor 200 according to the present invention. Phase-processor 200 converts an input signal $x(t)$ to an phase processed output signal $y(t)$ by altering the phase of various frequency components of $x(t)$ while leaving the amplitude of the signal in the various components substantially unchanged.

The output signal is generated by dividing the input signal into M components, each component matching the intensity of the signal in a specific frequency band. Apparatus 200 utilizes a plurality of band-pass filters 12 for this purpose. The signal in the i th frequency band is then phase-shifted by an amount ϕ_i utilizing a phase shifting network 14. As will be explained in more detail below, the specific ϕ_i values utilized will depend on the particular application in which audio processor 100 is being utilized. The ϕ_i are provided by controller 112 shown in Figure 1.

It is important that each of the band-pass filters preserve the phase of the

-8-

frequency component of $x(t)$ selected by the filter in question. The phase-shifted signals are then summed by signal adder 16 to form output signal $y(t)$.

5

Returning to Figure 1, the output of each phase-processor may be subjected to some form of post-processing. Hence, optional post-processing circuits 121-124 are shown in Figure 1. The post-processing in question may include amplifying the signals and/or mixing the signals with other signals derived from other sound sources. For example, additional stereophonic effects may be obtained by amplifying one channel relative to the remaining channels, thereby creating the illusion of a sound source closer to the speaker through which the corresponding output channel is played.

10

15 The output signal in the i th output channel will be denoted by $y_i(t)$. To simplify the following discussion, it will be assumed that there are only two output channels and that delay circuit 107 is utilized in the second output channel.

20

The cross-correlation measure of the output signals, $y_1(t)$ and $y_2(t)$ is determined by the phase shifts ϕ_i that were added to the various frequency components of $x(t)$. In the preferred embodiment of the present invention, the ϕ_i are chosen randomly between two limits which will be defined to be $P-\delta P$ and $P+\delta P$, respectively. Since $y_2(t)$ is merely $x(t)$ delayed by an amount to be discussed below, the ϕ_i are the phase difference between the $y_1(t)$ and $y_2(t)$ in the various frequency bands. Other methods for choosing the phase shifts will be described below.

25

30

The value of P (modulo 2π) determines the relative balance between the positive and negative peaks in the cross-correlation function. When P is equal to zero, the positive peak is at its maximum (close to 1) and the negative peak is at its minimum (close to 0). When P is equal to π , the positive peak is at its minimum (close to 0) and the negative peak is at its maximum (close to -1). When P is close to $\pi/2$ or $3\pi/2$, the positive and negative peaks are approximately of equal magnitude.

35

If a positive cross-correlation measure is to be obtained, then $-\pi/2 < P < \pi/2$. A negative cross-correlation measure is obtained when $\pi/2 < P < 3\pi/2$.

-9-

When P is approximately equal to $-\pi/2$ or $\pi/2$, the negative and positive peaks in the cross-correlation function are very close in magnitude and the cross-correlation measure could be positive or negative.

5

It has been found experimentally that P determines the image distance through the control of the ratio of the positive and negative peaks in the cross-correlation function. In loudspeaker reproduction, when $P = 0$, the image is close to the loudspeakers. As P increases from 0 to π the image moves closer to the listener. At values near to π , the image will appear to be close to the head, inside the head, or behind the head. As the value of P increases from π to 2π or as it decreases from π to 0 , the image moves back toward the loudspeakers.

10

The effect of P is approximately symmetrical about π , but not entirely. For $0 \leq P < \pi$, the positive peak in the cross-correlation function leads the negative peak. For $\pi \leq P < 2\pi$, the negative peaks leads the positive peak. Listeners report differences in the absolute distance of the sound source in these two conditions.

20

It may also be shown that δP determines the magnitude of the positive and/or negative peaks in the cross-correlation function. When δP is 0 , the magnitude of the peaks in cross-correlation function are at their maximum (close to $+/- 1$, but dependent on the value of P). As the value of δP increases from 0 to π , the magnitude of the peaks in the cross-correlation function decrease. When δP is equal to π , the magnitude of the peaks in the cross-correlation function are at their minimum (close to zero regardless of the value of P).

25

It is found experimentally that δP determines the perceived image width through control of the magnitude of the peaks in the cross-correlation function. In loudspeaker reproduction when $\delta P = 0$, the image is narrow and tightly focused. As δP increases from 0 to π the image becomes wider and more spatially diffuse. At values near to π , the image will appear to be extend from one speaker to the other. When δP is close to π , the magnitude of P ceases to have any substantial effect in controlling the apparent location of the image between the listener and the speakers. In this case, the sound is perceived as originating from a broad sound source located between the speakers.

30

35

-10-

Hence, the present invention may be utilized to control both the image width and distance. P is selected in order to provide the desired image distance. δP is selected in order to provide the desired image width. This may be accomplished by constructing a two-dimensional calibration curve for P as a function of image distance and δP as a function of image width, wherein the choice of P and δP are also dependent on each other.

The manner in which the phase shifts ϕ_i are chosen between the limits specified by P and δP is important in determining the quality of the output signals. In the preferred embodiment of the present invention, the ϕ_i are chosen by generating a sequence of random numbers between the limits in question.

Because of the finite number of frequency bands, it is found that different sets of random numbers produce slightly different effects. Hence, in the preferred embodiment of the present invention, a number of different sets of phase shifts are generated and the set producing the best effect, as judged by listening to the output signals, is selected.

Although the preferred embodiment of the present invention utilizes randomly selected phase shifts, other methods of choosing the phase shifts in question may be utilized without departing from the teachings of the present invention. Some of these methods are discussed below. In choosing a set of phase shifts within the range specified by P and δP , it is important that the phase shifts change direction frequently from band to band. Here, the phase shifts associated with two bands are said to change direction if the signal to the left speaker lags that to the right speaker in the first band while the signal to the left speaker leads that to the second speaker in the second band, or vice versa. As will be discussed in more detail below, this requirement is needed to prevent the perception of a "banded" or "broken" output signal. Consider three contiguous frequency bands having phase shifts ϕ_i , ϕ_{i+1} , and ϕ_{i+2} . On average, the change in phase shift should not be monotonic. That is, if $\phi_i > \phi_{i+1}$ then, on average, $\phi_{i+1} < \phi_{i+2}$. Similarly, if $\phi_i < \phi_{i+1}$ then, on average, $\phi_{i+1} > \phi_{i+2}$. Clearly, because of the random manner in which the phase shifts are chosen, there will be cases for which three consecutive phase shifts will be monotonic. However, on average this condition should be met.

-11-

To better understand the need for this requirement, consider the case in which one wishes to create the illusion of a physically broad sound source emitting sound along its surface between the two speakers. A sound component having a positive phase shift will be perceived as originating from a source which is closer to one speaker. A sound component having a negative phase shift will be perceived as originating from a source which is closer to the other speaker. The exact position at which each of the components is perceived will depend on the magnitude of the phase shift in question. Hence, the present invention produces a sound "image" that appears to emanate from a source that is made up of a collection of discrete sound components, each emitting sound in a specific frequency band and being located at a different position relative to the speakers. This requirement assures that, on average, signals from contiguous frequency bands will be perceived as originating from non-contiguous sources between the speakers.

The distribution of phase shifts will determine the spatial distribution of sound components. If the phase shift distribution is not uniform in phase, the spatial distribution will not be uniform in space. A uniform spatial distribution is desired since it is found experimentally that such a distribution remains uniform when the listener moves from the center line between the loudspeakers to a point off the center line. For example, when a listener is located left of the center line, sound from the left loudspeaker arrives before sound from the right loudspeaker which introduces a time delay in the arrival sound between the two ears. This time delay affects the phase difference at each frequency differently. A uniform distribution of phase provides the greatest assurance that that sound image is not altered by the time delay, since it results in another uniform distribution of phase.

The above discussion deals only with the phase shifts, ϕ_i . The manner in which the width of the bands is selected will now be discussed. If the bands are too broad, the listener will perceive a broken or banded image. However, if the bands are made too narrow, other problems are encountered.

As noted above, timbral shifts (so called coloration) of the output signal relative to input signal are to be avoided. These shifts arise from constructive and destructive interference. Such interference can arise from two independent

-12-

sources. First, the frequency bands into which the sound is divided have small overlaps. The degree of overlap depends on the specific filtering system used.

5 Consider such an overlap. The frequencies in the overlap region are contained in two adjacent bands. Each band has a different phase shift; hence, the overlap region will have components with different phase shifts at the same frequency. Depending upon the difference in phase shifts, there will be either constructive or destructive interference at the overlap frequencies when the signals from the

10 two bands are added back together after the phase shifting operation. This effect is minimized by choosing the broadest possible bands since the degree of overlap is relatively independent of the bandwidth.

15 The second source of interference will be referred to as spatial interference. When loudspeakers are utilized to reproduce the channels, the listener will receive overlapping sound fields, each field being generated by a different loudspeaker. At any given frequency, the signals from the two speakers will be perfectly correlated, since they differ only by a phase shift which depends on the frequency in question. Hence, there will be either constructive or destructive

20 interference between the signals depending upon the phase shift in question.

25 In addition, if the listener is not located on the center line between the speakers, there will be an additional phase shift added at each frequency. The additional phase shift results from the difference in distances between the listener and each speaker. For example, if the listener is closer to the right speaker, the signal from the left speaker will be delayed by a time equal to the difference in distance divided by the speed of sound. This time delay is equivalent to a frequency dependent phase shift being added to the output of one of the speakers. This added phase shift changes as the listener moves relative to the loudspeakers. Hence, at any given location relative to the loudspeakers, the listener is located in a sound field consisting of the sum of two signals having phase shifts which depend on the location of the listener and sound frequency. These signals will interfere with one another and produce a second timbral shift pattern which depends on the location of the listener.

30

35 It is known from psycho-acoustical research that there is a critical bandwidth below which the human ear can not discriminate. The critical bandwidth depends on frequency, varying from approximately 100 Hz at low frequencies

-13-

(< 2000 Hz) to approximately one seventh the center frequency of the band in question at high frequencies (> 2000 Hz).

5

Consider a band of critical bandwidth centered at a frequency F. If the frequency bands utilized in the present invention are much smaller than the critical bandwidth, then the critical frequency band in question will be made-up of a plurality of sub-bands, each with a different phase shift, ϕ_i . The intensity of the sound in the band will be the average of the intensities of each of the sub-bands.

10 Each sub-band will have an intensity which has been modified by the constructive or destructive interference resulting from the combining of the sound fields from the two speakers. This intensity will vary from 0 to 100 percent of the intensity that would have been present had the interference not taken place.

10

15 The undesired coloration results when the average intensity from band to band changes as a result of the interferences occurring at the sub-band level in each band. If the sub-bands were so small that there is a very large number of sub-bands in each band, then the change in average intensity from band to band would be negligible.

20

25 This may be seen as follows. The average intensity of each band is the average of the intensities of each sub-band. The intensity of each sub-band is reduced by a factor which is a function of a randomly selected variable, i.e., the ϕ_i . It is well known in the statistical arts that the standard deviation of the average of a function of a random variable for N values of the function goes to zero as N is increased to infinity. Thus, the variation form band to band is reduced as the number of sub-bands is increased.

25

30 Therefore, the coloration due to spatial interference of the sound waves produced by the left and right loudspeakers is minimized by reducing the bandwidth. As a result, one can not choose a bandwidth which simultaneously minimizes the timbral shifts from both factors.

30

35 In the preferred embodiment of the present invention, the bandwidth is chosen experimentally between about 50 Hz and twice the critical bandwidth. However, bandwidths as large as 4 critical bandwidths will function. If spatial interference coloration is small, then the larger bandwidth is found experimen-

-14-

tally to be more desirable. This will be the case when the listener is equidistant from the loudspeakers or wears headphones.

5

In addition, the question of an optimal bandwidth must be examined from the standpoint of sound material being processed. In the case of periodic and quasi-periodic tones such as speech and most musical instruments it is useful to organize the bands such that the harmonically related partials fall into separate bands. The fundamental will fall into a first band, the second harmonic into another, and so on. The limit to this rule is that bands not become smaller than a critical bandwidth since higher harmonics will naturally fall together into a single critical band. In the case of non-periodic or noise-like sounds, there is no fundamental. In this case, partials will likely fall into every adjacent band. It is useful that these bands be as small as possible and again that the phase of these adjacent bands shift rapidly. Experience has shown that the optimal bandwidth for non-periodic sounds is two critical bands wide.

We will now return to the issue of alternate methods of selecting phase shifts. The sound material being processed suggests different strategies. For periodic sounds, the non-adjacent bands containing harmonics should be phase shifted so that each partial is in a different spatial location. For non-periodic sounds, each adjacent band should be phase shifted so that adjacent bands of partials are in different spatial locations. Both strategies can be addressed together. Table 1 provides a list of center frequencies for bands and indications for left/right leading phase shifts such that adjacent bands lead in different directions and fundamental and second harmonics fall into non-adjacent bands leading in different directions up to the limit of critical band spacing. The left channel is defined to lead the right channel if $(\phi_R - \phi_L) > 0$. In the case in which a phase-shifted output signal is generated from the input signal, one of the ϕ 's will be zero. Hence, this is equivalent to requiring that the phase-shifts added to the frequency bands be chosen such that no three adjacent frequency bands are given phase shifts with the same sign.

-15-

	<u>Center Frequency of Bond</u>	<u>Left/Right</u>	<u>Direction of Phase Shift</u>
5	86	L	
	172	R	
	258	R	
	344	L	
	430	R	
10	516	L	
	602	R	
	689	R	
	775	L	
15	861	L	
	947	R	
	1033	R	
	1378	L	
	1636	L	
20	1894	R	
	2153	R	
	2411	L	
	2670	L	
	3186	R	
25	3617	R	
	4048	L	
	4737	L	
	5340	R	
	6201	R	
30	7235	L	
	8441	L	
	9819	R	
	11627	R	
	13781	L	
35	16537	L	
	22050	R	

Table 1

-16-

5 The exact phase shifts for each band can also be prescribed. For example, a vivid stereo separation for frequencies below 1,500 Hz may be achieved by selecting a phase value such as $+\pi$ (or $-\pi$) for each band. Alternatively, the phase shifts can be selected by choosing a random phase shift between 0 and $f\pi$ for the L bands and 0 and $-f\pi$ for the R bands, where f determines the apparent width of the image. If the image is to appear to emanate from a location between the speakers and the listener, a constant can be added to each phase shift in a manner analogous to that described above.

10 The above described embodiments of the present invention utilize band-pass filters and phase shift circuits. The same result may be obtained, however, by convolving $x(t)$ with a filter function $h(t)$ to produce $y(t)$. That is,

$$y(t) = \int x(t-z)h(z)dz \quad (2)$$

15 The transformation function $h(z)$ provides the phase shifting of the individual frequency bands.

20 The present invention preferably utilizes a digital input signal. If the signal source consists of an analog signal, it may be converted to digital form via a conventional analog-to-digital converter. In this case, each output signal consists of a sequence of digital values. The i th value for each output signal corresponds to the value of the output signal at a time iT , where T is the time between digital samples. In this case, the convolution operation given in Eq. (2) reduces to

$$30 \qquad y(nT) = y_n = \sum_m x_{n-m} h_m, \quad (3)$$

where m runs from 0 to $N-1$. The filter coefficients, h_m are calculated from

$$h_m = (1/N) \sum_k \exp(kmw + \phi_k) \quad (4)$$

35 Here, k runs from 0 to $N-1$, $w=2\pi/N$, $\exp(\alpha)=e^{i\alpha}$, and N is the total number of frequency samples.

-17-

5 In the above described preferred embodiment of the present invention, only one of the output signals is obtained from the input signal by processing the input signal, the other output signal being identical to the input signal. The output signal that is identical to the input signal can be delayed in time to compensate for the overall delay introduced by the processing. In the case that the processing is performed by convolution, this delay will be approximately equal to half the length of the convolution sequence.

10 In the preferred embodiment, the cross-correlation measure value is determined by the relationship of the processed output channel to the unprocessed output channel. That is, one of the output channels is not phase-processed. It is found experimentally that the presence of an unprocessed channel reduces the perceived effect of any small timbral shifts. Those skilled in the art will also recognize that the same interchannel relationship can be achieved in an implementation in which both output signals are processed. In such an implementation, the phase characteristics we have described for the processed signal in the preferred embodiment are implemented such that the interchannel phase differences satisfy the conditions in question.

15 Although the above embodiments of the present invention have been described with reference to stereophonic output signals, it will apparent to those skilled in the art that the principles described above may be utilized for providing more than two output signals. For example, in theatrical sound systems four or more output channels are often utilized. Each of the output channels can be processed by a phase-processor according to the present invention. Each phase-processor would utilize its own set of phase shifts, ϕ_i . Each such set of phase shifts would be different from those used by other said phase-processors.

20 Unlike prior art systems, the perceptual effects obtained with the present invention are resilient in loudspeaker reproduction, even when the listeners are far off the line equidistant between the two loudspeakers and even when the reproduction environment is reverberant. Experiments have shown that the effect is present even when the distance between the listener and each of the loudspeakers differs by as much as 15 meters in typical reproduction settings.

25 The output signals provided by the present invention may be played

-18-

through conventional speakers or headphones. These signals may also be recorded onto conventional stereophonic recording media for subsequent playback through conventional stereophonic equipment. Such an audio recording has at least two channels. When the final mixing is completed, each sound track can be viewed as being composed of two signals $Q(t)$ and $R(t)$. The $Q(t)$ signal is the result of the processing by the present invention. The $R(t)$ signal represents other processing such as mixing channels of information which are not processed by the present invention.

Consider a recording having two channels. The first sound track would be composed of $Q_1(t)$ and $R_1(t)$ and the second sound track would be composed of $Q_2(t)$ and $R_2(t)$. The amplitudes of the signals $Q_1(t)$ and $Q_2(t)$ at any given frequency will be denoted by $A_1(f)$ and $A_2(f)$. In a recording according to the present invention, $A_1(f) = gA_2(f)$ where g is a gain related constant. The phase of $Q_1(t)$ at any given frequency f will differ from that of $Q_2(t)$ by an amount $\phi(f)$, where $\phi(f)$ varies between $P-\delta P$ and $P+\delta P$, and $\phi(f)$ is a rapidly changing function of frequency.

For the purposes of this discussion, $\phi(f)$ is defined to be rapidly varying if the following criteria are met. Consider the frequency spectrum as being broken into bands of width no larger than four critical bandwidths. Consider the average value of $\phi(f)$ in any given band. $\phi(f)$ is said to be a rapidly varying function of f if, on average, the sign of the difference in the average value of $\phi(f)$ between bands is zero. If this criterion is met, then, on average, adjacent frequency bands will lead through different speakers when P is 0.

There has been described herein a novel audio processing method and apparatus. Various modifications to the present invention will become apparent to those skilled in the art from the foregoing description and accompanying drawings. Accordingly, the present invention is to be limited solely by the scope of the following claims.

-19-

WHAT IS CLAIMED IS:

- 5 1. An audio processing system for generating a plurality of output channel signals from an input sound signal, said system comprising:
means for receiving said input sound signal;
- 10 phase-processing means for generating a phase-processed signal comprising a signal which is substantially equal to the sum of M band-limited signals, the ith said band-limited signal having an amplitude substantially equal to that of said input sound signal in a predetermined frequency range $f_i \pm \delta f_i$ and a phase which differs from the phase of said input sound signal in said predetermined frequency range by an amount ϕ_i , i running from 1 to M, wherein $M > 2$ and ϕ_i is chosen between $P - \delta P$ and $P + \delta P$, wherein ϕ_i is a rapidly varying function of i; and
15 means for generating one of said channel signals from said phase-processed signal.
- 20 2. The system of Claim 1 wherein said δf_i are chosen such that δf_i is less than twice the critical bandwidth at f_i for all values of i for which f_i is less than some predetermined frequency.
- 25 3. The system of Claim 1 wherein said δf_i are chosen such that δf_i is substantially equal to the critical bandwidth at f_i for all values of i for which f_i is less than some predetermined frequency.
- 30 4. The system of Claim 1 wherein said f_i and δf_i are chosen such that harmonically related partials are in different said frequency ranges for frequencies below a predetermined frequency.
- 35 5. The system of Claim 1 further comprising delay means for generating a delayed signal having an amplitude and phase substantially equal to that of said input signal; and
means for generating one of said channel signals from said delayed signal,

-20-

said generated channel signal being different from the channel signal generated from said phase-processed signal.

5

6. A method for processing an input sound signal to generate a plurality of output channel signals, said method comprising the steps of:

receiving said input sound signal;

10

generating a phase-processed signal which is substantially equal to the sum of M band-limited signals, the ith said band-limited signal having an amplitude substantially equal to that of said selected input sound signal in a predetermined frequency range $f_i \pm \delta f_i$ and a phase which differs from the phase of said input signal in said predetermined frequency range by an amount ϕ_i , i running from 1 to M, wherein $M > 2$ and ϕ_i is chosen between $P - \delta P$ and $P + \delta P$, wherein ϕ_i is a rapidly varying function of i; and

15

generating one of said channel signals from said phase-processed signal.

20

7. The method of 6 wherein said δf_i are chosen such that δf_i is less than twice the critical bandwidth at f_i for all values of i for which f_i is less than some predetermined frequency.

25

8. The method of 6 wherein said δf_i are chosen such that δf_i is substantially equal to the critical bandwidth at f_i for all values of i for which f_i is less than some predetermined frequency.

30

9. The method of 6 wherein said f_i and δf_i are chosen such that harmonically related partials are in different frequency ranges for frequencies below a predetermined frequency.

35

10. The method of 6 further comprising the step of generating a delayed signal from input sound signal, and generating one of said channel signals from said delayed signal, said generated channel signal being different from the channel signal generated from said phase-shifted signal;

11. An audio recording comprising first and second channels, said first

-21-

channel comprising a signal which is the sum of two signals, $Q_1(t)$ and $R_1(t)$ and
said second channel comprising the sum of two signals $Q_2(t)$ and $R_2(t)$, wherein
5 $A_1(f) = gA_2(f)$, $A_1(t)$ being the intensity of $Q_1(t)$ at frequency f , $A_2(t)$ being the
intensity of $Q_2(t)$ at frequency f , and g being a constant, the phase of $Q_1(t)$ at
any given frequency f differs from that of $Q_2(t)$ by an amount $\phi(f)$, $\phi(f)$ varying
between $P - \delta P$ and $P + \delta P$, and $\phi(f)$ being a rapidly changing function of frequen-
cy f .

10

15

20

25

30

35

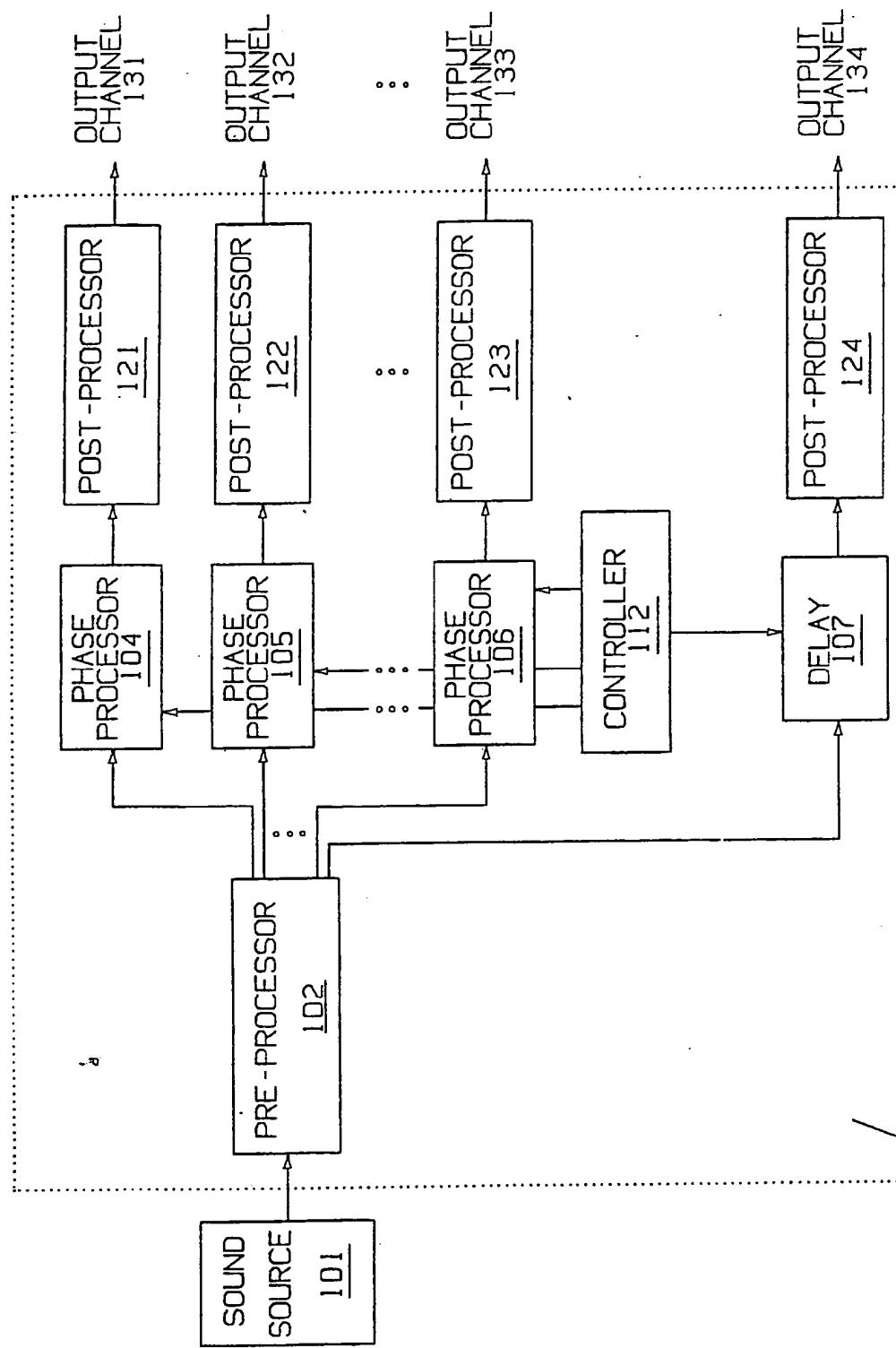


FIGURE 1

100

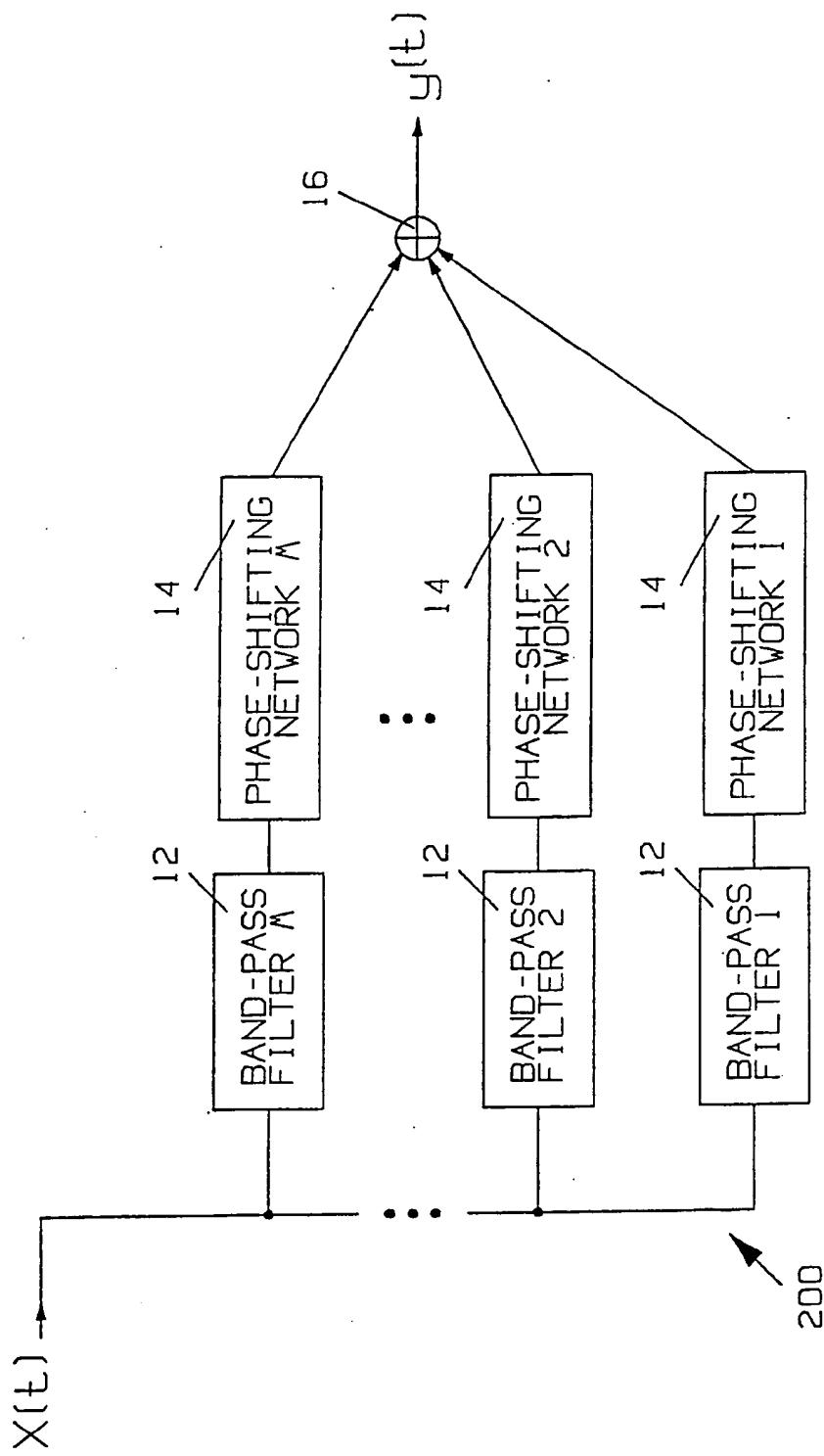


FIGURE 2

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US91/04166

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ⁶

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(5): H04S 1/00

US CL : 381/1

II. FIELDS SEARCHED

Minimum Documentation Searched ⁷

Classification System	Classification Symbols
U.S.	381/1,17,97

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched ⁸

III. DOCUMENTS CONSIDERED TO BE RELEVANT ⁹

Category ¹⁰	Citation of Document, ¹¹ with indication, where appropriate, of the relevant passages ¹²	Relevant to Claim No. ¹³
X Y	JP, A, 58-190 199 (KOVICHI KUROEUMI) 07 NOVEMBER 1983 See english abstract ("Purpose" and "Constitution") and figures 1,3, and 5.	1,5,6,10 2-4,7-9,11
Y	US, A, 3,670,106 (ORBAN) 13 JUNE 1972 See column 2, lines 47-50.	2-4,7-9
Y	US, A, 4,908,858 (OHNO) 13 MARCH 1990 See figure 1 and abstract.	11

- Special categories of cited documents: ¹⁰
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search

13 AUGUST 1991

International Searching Authority

ISA/US

Date of Mailing of this International Search Report

13 SEP 1991

Signature of Authorized Officer

Nguyen Ngoc-Ho
F.W. ISEN

NGUYEN NGOC-HO
INTERNATIONAL DIVISION